# The three heavy-chain precursors for the inter-α-inhibitor family in mouse: new members of the multicopper oxidase protein group with differential transcription in liver and brain

Philippe CHAN,* Jean-Loup RISLER,† Gilda RAGUENEZ* and Jean-Philippe SALIER*‡

*Institut National de la Santé et la Recherche Médicale, Unit 78, and Institut Fédératif de Recherches Multidisciplinaires sur les Peptides, Boisguillaume, France and †Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, Gif/Yvette, France

The inter-α-inhibitor (IαI) family is comprised of the plasma protease inhibitors IαI, inter-α-like inhibitor (IαLI), pre-α-inhibitor (PαI) and bikunin. IαI, IαLI and PαI are distinct assemblies of bikunin with one of three heavy (H) chains designated H1, H2 and H3. These H chains and bikunin are respectively encoded by a set of three *H* genes and an $\alpha_1$-microglobulin/bikunin precursor (AMBP) gene. All four gene products undergo maturation steps from precursor polypeptides. The full-length cDNAs for the H1-, H2- and H3-chain precursors were cloned from a mouse liver cDNA library and sequenced. Extensive searches of amino acid sequence similarities to other proteins in databanks revealed (i) a highly significant similarity of the C-terminal sequence in the three H-chain precursors to the multicopper-binding domain in the group of multicopper oxidase proteins and (ii) the presence of von Willebrand type-A domains in the mature H chains. Amino acid sequence comparisons between the three mouse H1-, H2- and H3-chain precursors and their human counterparts allowed us to appraise the timing and order of occurrence of the three H-chain genes from a shared ancestor during mammalian evolution. Owing to a multiple alignment of the six mouse and human nucleotide sequences for these H-chain precursors, a reverse transcriptase PCR assay with degenerate oligonucleotides was designed, allowing us to (i) present evidence that no mRNAs for further *H* genes exist in mouse liver and (ii) demonstrate a previously undescribed transcription of the H2- and H3-chain mRNAs in mouse brain, which contrasts with the expression of all four, H1, H2, H3 and AMBP, mRNAs in liver.

## INTRODUCTION

The inter-α-inhibitor (IαI) family (reviewed in refs. [1,2]) encompasses a set of at least four mammalian plasma protease inhibitors, designated IαI, inter-α-like inhibitor (IαLI), pre-α-inhibitor (PαI) and bikunin. These molecules are exclusively synthesized in liver. They are made up of a series of four distinct polypeptides, including three evolutionarily related heavy (H) chains H1, H2 and H3, and a light chain designated bikunin after its tandem arrangement of Kunitz-type protease inhibitor domains. IαI corresponds to an (H1 + H2 + bikunin) assembly, IαLI is comprised of (H2 + bikunin) and PαI is made up of (H3 + bikunin); bikunin is also present as a free molecule in plasma [1–5]. The bikunin polypeptide originates from an $\alpha_1$-microglobulin/bikunin precursor (AMBP) [6]. This precursor undergoes post-translational processing which results in two mature polypeptides, i.e. $\alpha_1$-microglobulin and bikunin, only the latter being a component of the various molecules found in the IαI family. Likewise, the mature H1, H2 and H3 chains originate from three H1, H2 and H3 precursors. These precursors undergo processing steps that trim a short N-terminal propeptide sequence as well as a stretch of about 250 amino acid residues at the C-terminus [7,8]. Ultimately, the series of IαI family molecules are made up of the mature H chains and bikunin held together via quite unusual covalent protein–glycosaminoglycan–protein cross-links [4,9]. Contrasting with these detailed features of chain structures and assemblies in the family, and the protease inhibitory activities of IαI family members as observed *in vitro*, very little is known about the functions of these molecules *in vivo*. One or another among the IαI family members could possibly act as an endothelial cell growth factor [10], have a hyaluronan-binding capacity [11,12], or participate in the inflammatory response to an acute infection [5].

Using mouse as a model, we are currently interested in determining precisely the biological functions and regulation of expression of bikunin and the H1–H3 chains in the IαI family. We have demonstrated that, as in man, four *AMBP*, *H1*, *H2* and *H3* genes are present in mouse. These genes display arrangements, chromosomal locations, transcript sizes and a liver-restricted expression that are quite homologous in the two species [13–16]. We now present the sequences of the full-length cDNAs for mouse H1-, H2- and H3-chain precursors. The deduced protein sequences allow us to infer previously unsuspected functions for these H chains, as well as to date the divergence of this multigene family during mammalian evolution. These cDNA sequences were further used to design a reverse transcriptase PCR (RT-PCR) assay. This has allowed us to demonstrate a previously unreported transcription of the H2- and H3-chain mRNAs in mouse brain, which contrasts with the known expression of all four, H1, H2, H3 and AMBP, mRNAs in liver.

## MATERIALS AND METHODS

### Probes, library screening and nucleotide sequencing

A mouse (strain C56BLG/N) liver cDNA library cloned at the *Eco* RI site in the λgt11 phage vector was initially prepared in Dr. F. Gonzalez's laboratory (National Institutes of Health, Bethesda, MD, U.S.A.) and further amplified by Professor P. Arnaud (Medical University of South Carolina, Charleston,

Abbreviations used: AMBP, $\alpha_1$-microglobulin/bikunin precursor; H, heavy; IαI, inter-α-inhibitor; IαLI, inter-α-like inhibitor; ORF, open reading frame; PαI, pre-α-inhibitor; RT-PCR, reverse transcriptase PCR.

The nucleotide sequence data provided in this paper have been deposited in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession numbers X70391, X70392 and X70393.

* To whom correspondence should be addressed, at INSERM U-78, B.P. 73, 76233 Boisguillaume Cedex, France.

SC, U.S.A.). The library was propagated in *Escherichia coli* Y1090 strain and the positive colonies were plaque-purified by successive rounds of nylon filter hybridization. Human partial H1-, H2- and H3-chain cDNA probes cloned from a liver cDNA library [13,17,18] or newly isolated mouse cDNA probes were labelled by the random oligonucleotide procedure with [$\alpha$-$^{32}$P]dCTP (3000 Ci/mmol; Amersham) at a specific radioactivity of $0.5 \times 10^9$–$2 \times 10^9$ c.p.m./$\mu$g. Hybridizations and post-hybridization washes were performed at high (rodent cDNA probes) or low (human cDNA probes) stringency as previously described [5,13]. Autoradiographic exposure to Hyperfilm-MP (Amersham) was performed at −80 °C with intensifying screens. *Eco*RI-digested inserts of the positive λgt11 clones were subcloned into a pUCBM20 or pUCBM21 plasmid (Boehringer-Mannheim). Dideoxy sequencing at both ends of the insert was carried out with alkali-denatured double-stranded DNA, a universal (forward or reverse) primer for pUC-based plasmids, [$\alpha$-[$^{35}$S]thio]dATP (1000 Ci/mmol; Amersham) and a T7 polymerase sequencing kit (Pharmacia). For every clone detailed below the complete sequence of the insert was obtained on both strands from a series of nested deletions made with an exonuclease III-based kit (Pharmacia). Compression artifacts were resolved with 7-deaza-dGTP (Pharmacia kit) and/or H-chain-specific oligonucleotides. Direct dideoxy sequencing of some RT-PCR products (see below) without subcloning was carried out as described [19] using a specially designed degenerate oligonucleotide as a sequencing primer (see the Results section). Finally, some RT-PCR products were first subcloned into a pGEM-T or pUCBM20 plasmid (see below) before conventional sequencing. The assembly of partial nucleotide sequences and some data analyses were performed using DNASIS software (Pharmacia).

### Computerized searches for amino acid sequence similarities in protein databanks

Similarities between one or more H-chain sequences and the series of polypeptide sequences currently available in various databanks were extensively investigated as follows. First, about 30 short blocks of nine or more highly conserved amino acid residues between H chains, as indicated by a multiple sequence alignment (see the Results section), were separately matched against the SWISS-PROT databank (release 22) with the EMBLScan software, and the threshold for a significant similarity between polypeptides was arbitrarily chosen as 50% or higher identity over at least ten residues; the possible similarities thus found were further examined by a complete chain alignment. Second, the cDNA-deduced complete amino acid sequences of all three mouse H chains were compared with the PIR/MIPS databank (release 35) using the FASTA program [20] and with the non-redundant protein databank at NCBI using the BLAST program [21]. In addition, the program CLUSTAL [22] was used to build up a multiple alignment of the six human and mouse H chains. This multiple alignment was used as an input to build a profile [23] which in turn became a query sequence for searching the PIR/MIPS databank with the program PROFILESEARCH from the GCG package [24] running on our local VAX/VMS 6300 computer. Each of the human and mouse sequences was also compared with the PROSITE (release 9) [25] and BLOCKS [26] databanks. All subsequent pairwise alignments were performed with the BESTFIT program from GCG.

### PCR and RT-PCR

Total RNAs from mouse tissues were prepared by the guanidinium isothiocyanate procedure followed by centrifugation onto a CsCl cushion. RNA integrity was controlled by visual inspection of the 18 S and 28 S rRNA bands in agarose gels. RNA concentrations were determined from the absorbance at 260 nm. RT-PCR with total RNAs or conventional PCR with cloned DNA in plasmids used as controls were carried out with a pair of specific oligonucleotides (Genosys) at 20 pmol each, a 10 × buffer (Promega), dNTPs (Pharmacia) 200 $\mu$M each, 2.5 mM $MgCl_2$ and 1.25 units of *Taq* polymerase (Promega), in a 100 $\mu$l final reaction volume. The PCR conditions used for a pair of H-chain-specific degenerate oligonucletides as well as a pair of AMBP-specific oligonucleotides (sequences detailed in the legends to Figures 5 and 6) were: denaturation step for 4 min at 95 °C; five cycles (94 °C for 20 s/60 °C for 1 min/72 °C for 2 min) with ramping (6 s/ °C, from 60 to 72 °C); 25 cycles as above without ramping; last elongation step at 72 °C for 10 min. Before standard PCR steps, the RT-PCR included a cDNA first-strand synthesis carried out in 30 $\mu$l with 2 $\mu$g of total RNAs, 60 units of RNAsine (Promega), dNTPs (1 mM each), 250 pmol of random hexamer primers (dpN6 from Pharmacia) and 400 units of RT from murine moloney virus (Gibco–BRL) in 10 mM Tris/HCl, pH 8.3, buffer containing 15 mM KCl, 0.6 mM $MgCl_2$ and 25 $\mu$M dithiothreitol. After completion of the reaction at 37 °C for 1 h, 15 $\mu$l of the reaction products was used for the RT-PCR step detailed above. Control experiments for RT-PCR included template digestion with either RNAase-free DNAase (from Gibco–BRL) or DNAase-free RNAase A (heated at 80 °C for 2 h before use). A control plasmid was made of the full-length mouse H3-chain cDNA (see the Results section) subcloned at the *Eco*RI site of the polylinker into the pUCBM21 plasmid (Boehringer-Mannheim). Finally, the gel-purified RT-PCR products were subcloned into a pGEM-T plasmid (Promega) or a *Sma*I-restricted pUCBM20 plasmid; some restriction fragments from the RT-PCR products were subcloned into a (*Sma*I + *Kpn*I)- or a (*Sma*I + *Sal*I)-restricted pUCBM20 plasmid.

## RESULTS

### Nucleotide and deduced amino acid sequences of the H1-, H2- and H3-chain cDNAs in mouse

From a mouse liver cDNA library, a series of cDNA clones were identified for the H1-, H2- and H3-chain precursors. They were subcloned into a pUCBM20 or pUCBM21 plasmid and sequenced on both strands by a nested deletion strategy. As shown in Figure 1, two H1 or three H2 clones were used to cover the entire H1- or H2-chain cDNA sequence respectively. They were designated λMoH1–7 and λMoH1-9 (H1-chain cDNA) or λMoH2-14, λMoH2-18 and λMoH2-19 (H2-chain cDNA). A single H3 clone, designated λMoH3-38, provided the full-length sequence for the H3-chain cDNA. The resulting nucleotide sequences of the three full-length cDNAs along with the three deduced open reading frames (ORFs) are provided separately (EMBL/GenBank/DDBJ Nucleotide Sequence Databases: X70391, X70392, X70393).

The H1-chain cDNA is 2904 bp long. The largest ORF is 907 amino acids long and starts with a leucine-rich signal peptide located from $Met^{-26}$ to $Gly^{-1}$ (the signal peptides are numbered negatively). The proposed borders for this signal peptide best respect the expected hydropathy profile and the so-called '−3/−1 rule' [27]. The H2-chain cDNA is 3125 bp long. The largest ORF starts with two methionines in neighbouring positions (first and fifth amino acids). None of these is encoded by an AUG initiation codon with a nucleotide environment that satisfactorily fits Kozak's minimum consensus [28]. Therefore

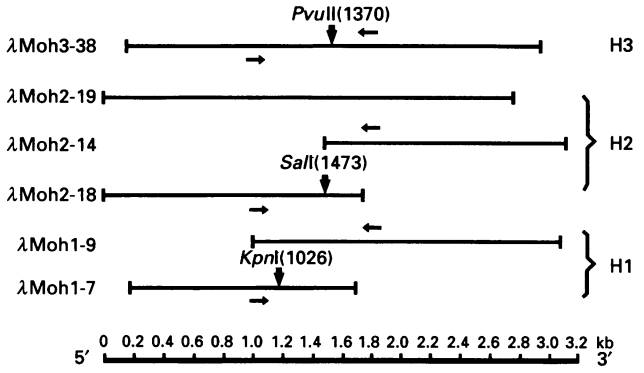


**Figure 1 Mouse cDNAs for the H1-, H2- and H3-chain precursors: cDNA clones; location and restriction mapping of RT-PCR products**

A code for the λgt11 clones is indicated on the left and the assignment of each clone to an H-chain precursor is quoted on the right. The horizontal arrows in a forward or reverse orientation correspond to sequences that are highly homologous between the three H-chain cDNAs and were used to design degenerate, sense or antisense oligonucleotides for RT-PCR. The restriction sites used as controls for identification of the RT-PCR products are indicated with vertical arrows (numbering according to the corresponding mouse cDNA sequence provided in the EMBL data library).

either of these AUG codons may be the genuine translation-start site. When taking the most upstream AUG into account, the deduced peptide sequence is 950 amino acids long and starts with a leucine-rich signal peptide most likely located from $Met^{-22}$ to $Ala^{-1}$, as inferred from ref. [27]. The H3-chain cDNA is 2775 bp long. The largest ORF is 886 amino acids long and starts with a leucine-rich signal peptide most likely located from $Met^{-18}$ to $Gly^{-1}$, as inferred from ref. [27].

## Homologies and differences between the H-chain precursors

Amino acid sequence comparisons were made between the three H-chain precursors as well as between mouse and human sequences [7,17,29,30]. The resulting sequence alignment of the six H1, H2 and H3 polypeptides from human and mouse is presented in Figure 2. The levels of strictly conserved amino acids (Figure 3) indicate that all three H-chain sequences have been highly conserved from primates to rodents, the identical residues between species for a given H-chain precursor being 83% or more. Furthermore, in both human and mouse the sequence homology is strongest between H3 and H1 whereas the H2 sequence is most distantly related (Figure 3). Also, and excluding the signal peptides from our analysis, we note that all cysteine residues in the H-chain precursors are conserved from mouse to human (Figure 2) with the exception of the $Cys^{81}$ residue in the mouse H3 chain which is replaced by $Ser^{81}$ in the human counterpart. Finally, it is noteworthy that the potential N-glycosylation sites found in human H-chain precursors are all strictly conserved in mouse (see Figure 2: $Asn^{307}$, $Asn^{611}$, $Asn^{787}$ for H1; $Asn^{98}$, $Asn^{120}$, $Asn^{449}$ for H2; $Asn^{120}$, $Asn^{611}$ for H3) whereas the latter species has gained one further site in each of these H chains (H1: $Asn^{310}$; H2: $Asn^{267}$; H3: $Asn^{757}$).

The comparison between H1-, H2- and H3-chain precursors also provides important features, regardless of the species. The overall sequences in the precursors (Figure 2) can be grossly divided into three segments: (i) a first half (residues 69–527) where strictly conserved (reverse printed) residues are evenly distributed along the sequence; (ii) a further 20% (residues 528–713) where blocks of conserved residues alternate with long stretches of quite divergent sequences including many alignment gaps and even rodent-associated deletions; (iii) a last 30% (residues 714–954) the arrangement of which is similar to the first half. Therefore one may predict that some differences in functions between the three H chains would result in particular from the intermediate segment in the precursors, i.e. the C-terminal area of the mature chains. Finally, the sequence $DPHFII^{713}$ is strictly conserved between chains and species; notably it contains the $Asp^{708}$ residue which is the C-terminus of the mature H2 and H3 chains in human [4,9], strongly suggesting that this homologous $Asp^{708}$ residue is the C-terminus of all three mature chains in human and mouse.

## Presence of von Willebrand type-A domains and multicopper oxidase domains in the H chains

To discover some potential biological functions of the H chains, an extensive search for amino acid sequence similarities between the H-chain precursors and other polypeptides was carried out with help from several, non-redundant databanks and computer programs (see the Materials and methods section). First, the trivial observation that the H-chain sequences contain amino acid stretches or domains with such potential activities as amidation or myristoylation sites and ck2- or pkc-associated phosphorylation sites was made during our search with the PROSITE databank [25] and is not detailed here. The major findings of our searches are presented in Table 1 and the amino acid sequences involved in the three H-chain precursors are underlined in Figure 2. Significant similarities were found between one or more H-chain segment(s) and (i) a tyrosine phosphorylation site, (ii) two proteins containing von Willebrand type-A domains, and (iii) the copper-binding domain of multicopper oxidase proteins. As shown in Figure 4, a sequence that fits the consensus provided in ref. [25] for this multicopper oxidase domain is (i) unambiguously present in the mouse H1- and H3-chain precursors, within the C-terminal sequence which is trimmed off in the mature polypeptides and (ii) present, with a lower similarity level, in the homologous area of the mouse H2-chain precursor. Identical observations apply for the human chains (compare human and mouse sequences in Figure 2).

## A new RT-PCR assay for any H-chain-related transcripts detects only three transcripts in mouse liver

As the three H1-, H2- and H3-chain cDNAs correspond to evolutionarily related genes that originate from a shared ancestor (see the Discussion), we wondered whether more than three genes and mRNAs could exist and we sought further H-chain mRNAs in mouse liver by using a specially designed analysis by RT-PCR. The three mouse cDNA sequences were compared with the human cDNAs by a multiple alignment (not shown) that highlighted two regions of strongly conserved nucleotide sequences. These regions allowed us to design a pair of degenerate oligonucleotides (sequences in the legend to Figure 5) that were used in RT-PCR. The locations of the three sequences in mouse H-chain cDNAs used to design this pair of oligonucleotides are indicated in Figure 1 (arrows). Furthermore, the choice of this oligonucleotide pair was made so as to amplify an H1-, H2- or H3-chain sequence (830 bp in length) harbouring a unique, H1-, H2- or H3-chain-specific restriction site, namely *Kpn*I, *Sal*I or *Pvu*II respectively, as depicted in Figure 1. The relevant re-

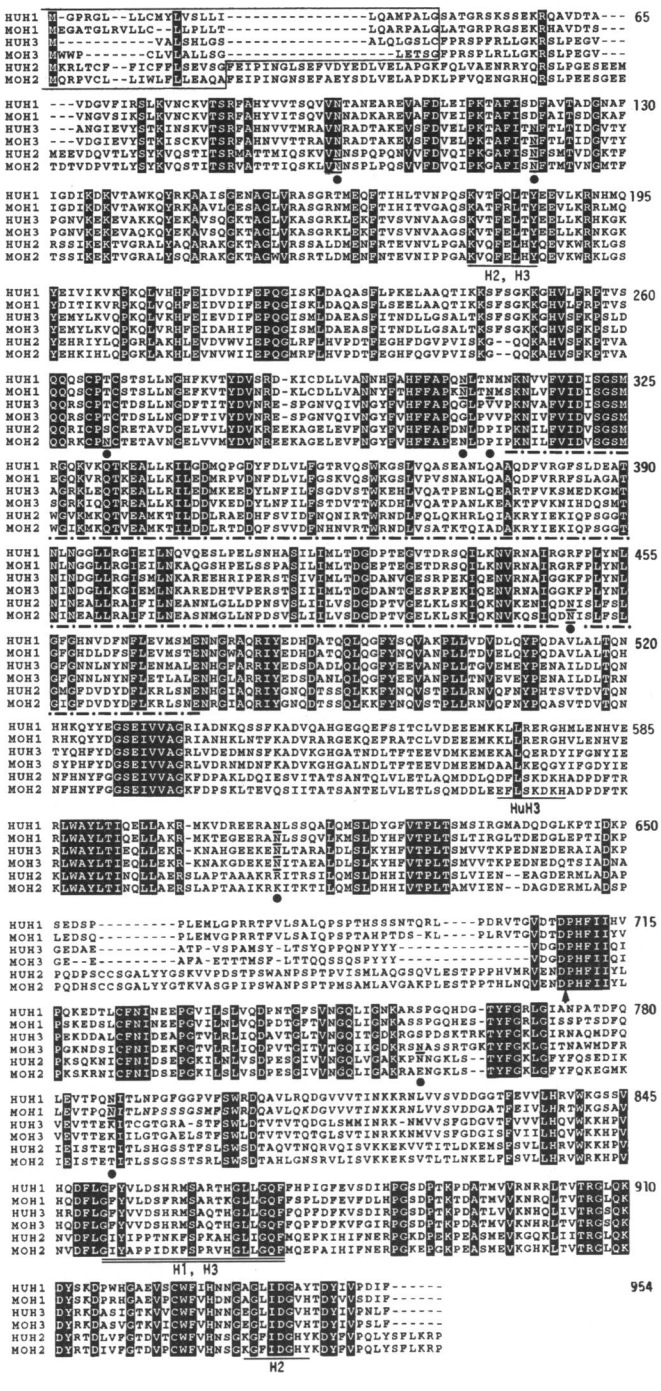

**Figure 2 Multiple sequence alignment for the human and mouse H1-, H2- and H3-chain precursors: indications of domains with potential activity**

Each human (HU) or mouse (MO) precursor is identified on the left. The residues that are identical in all six chains are reverse printed. Note that a very high level of conservative changes can be observed between chains but this is not highlighted for clarity. Gaps (—) have been introduced for maximum similarity. Amino acid numbering is indicated on the right; this arbitrary numbering is specific to this multiple alignment and does not match any of the discrete numberings otherwise used for the human [7,17,29,30] and mouse H-chain precursors. The first four residues in the human H1 signal peptide (MDGA [29]) or mouse H2 signal peptide (MSSK in this study) have been omitted: they are not found in the mouse H1 precursor (this study) or human H2 precursor [7] respectively. The most likely borders of the signal peptides (see the text) are boxed. Columns containing one or more potential N-glycosylation sites are indicated with a closed circle, and the involved N residues in the column are underlined (some underlinings are reverse printed). The probable C-terminal cleavage point shared by the mature H1, H2 and H3 chains is marked with an arrowhead. The matches of some H-chain domains with other polypeptide sequences suggesting potential biological activity (see Table 2) are

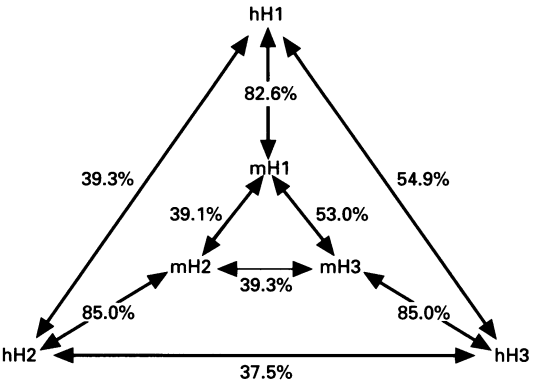


**Figure 3 Amino acid identities in the three H1-, H2- and H3-chain precursors: comparisons between chains and between species**

The number of identical amino acids shared by a pair of polypeptides is given as a percentage of the total number of residues in the longest polypeptide compared in each pair. Conservative amino acid changes were not taken into account. The signal peptide sequences were excluded from these calculations. The cDNA-deduced human sequences are from refs. [7,29,30]. h, human; m, mouse.

striction site was used to characterize the presence of the corresponding RT-PCR-amplified H1-, H2- or H3-chain mRNA within the bulk of amplified H-chain mRNA-related products.

The results obtained for adult mouse liver RNAs are presented in Figure 5. Although the amplified H1-, H2- and H3-chain products were all expected to be 830 bp in length, the crude RT-PCR product was reproducibly made of two distinct bands, 830 and 870 bp in size (lane 1). Digesting the bulk of RT-PCR products with *Kpn*I (lane 3) or *Pvu*II (lane 5) left the 870 bp band intact whereas this band was completely digested by *Sal*I (lane 4). This restriction pattern as well as the sequencing of these bands (Table 2, rows 1 and 2) unambiguously identify the unexpected 870 bp band as the H2-chain RT-PCR product whereas the 830 bp band represents a mixture of H1- and H3-chain products. We suspect that the abnormally slow migration of the H2-chain RT-PCR band probably results from a sequence-specific bend in the amplified H2 segment. Finally, the expected three pairs of restriction fragments were obtained from the RT-PCR products digested with *Kpn*I, *Sal*I and *Pvu*II used in conjunction (Figure 5, lane 2) or separately (lanes 3–5). The grossly similar amounts of these three pairs of restriction fragments (lane 2) indicate similar levels of mRNA for three H-chain precursors in adult mouse liver.

The data above demonstrate that our degenerate oligonucleotides and RT-PCR assay are able to co-amplify various H-chain-related mRNAs. After digestion with all three enzymes at once, trace amounts of undigested RT-PCR product were still visible (Figure 5, lane 2), suggesting the existence of previously undescribed H-chain mRNAs. Subcloning and partial sequencing of this undigested product provided the known sequences for the H1-, H2- and H3-chain cDNAs (not shown). This indicates that further H-chain mRNA lacking all three restriction sites does not account for this undigested product, which is probably explained by *Taq* polymerase-induced errors at the restriction sites in some

indicated as follows: potential tyrosine phosphorylation sites are underlined; a potential multicopper oxidase domain is double underlined; a von Willebrand type-A domain is marked with a broken underline. When not all H-chain precursors in the multiple alignment match the considered sequence, those human and/or mouse H chains that do match it are quoted below the underlining.

**Table 1 Potential biological activities associated with H-chain precursors, as indicated by sequence similarities to other proteins**

| Databank | Software used for search | Motif used for search* | Result† | Sequence similarity and associated activity if any‡ |
|---|---|---|---|---|
| PROSITE | FASTA | Hu/Mo/H1,2,3 | H2 and H3 | Tyrosine phosphorylation site |
| | | | H1 and H3 | Multicopper oxidase, $Ca^{2+}$-binding domain § |
| PIR/MIPS | FASTA + BESTFIT | Hu/Mo/H1,2,3 | MoH1, MoH3, and Hu/MoH2 | Dihydropyridine $Ca^{2+}$ channel, α2 subunit (human/rodent)‖ |
| PIR/MIPS | BLAST | Hu/Mo/H1,2,3 | HuH2 | Dihydropyridine $Ca^{2+}$ channel, α2 subunit (rabbit/rat)‖ |
| | | | MoH1 | Collagen VI, α3 chain (chicken cornea)‖ |

* A search for a match in the databank was made with each H-chain sequence (Hu, human; Mo, mouse) separately. Other kinds of searches (see the Materials and methods section) gave negative results.

† The indicated chain(s) contains a sequence that is similar to sequence(s) in the databank. Unless further specified, the similarity found is conserved in human and mouse.

‡ The corresponding sequences within the H-chain precursor(s) are outlined in Figure 3.

§ Similarity further detailed in Figure 4.

‖ Proteins with von Willebrand type-A domain.

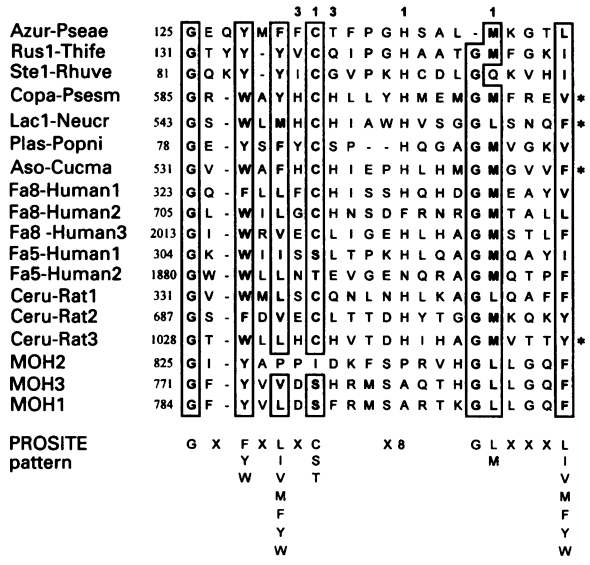


**Figure 4 Multiple alignment of the multicopper oxidase domain found in the mouse H-chain precursors and other proteins**

A consensus for proteins that belong to the group of multicopper oxidases with a multicopper oxidase domain of type 1 was provided by the PROSITE databank [25] and is shown at the bottom: prosite pattern (X, undefined amino acid residue). This multicopper oxidase domain of type 1 does not make any assumption on the presence of genuine copper-binding residues and thus can detect domains that have lost the ability to bind copper [25]. Some proteins with a proven copper-binding capacity are marked with an asterisk. The positions at which copper ligands occur in some of the domains are numbered at the top according to the type of copper binding [33]. Identical amino acid residues found in all sequences are boxed. The number on the left of each sequence refers to the position of the reverse-printed glycine residue in the original sequence. Each protein is identified on the left: MOH1, MOH2 or MOH3, mouse H1-, H2- or H3-chain precursor. The other proteins listed are from the SWISS-PROT databank and contain one or more identified sites with a proven or suspected copper-binding activity (reviewed in ref. [33]): Azur-Pseae, azurin from *Pseudomonas aeruginosa*; Rus1-Thife, rusticyanin from *Thiobacillus ferro-oxidans*; Ste1-Rhuve, stellacyanin from *Rhus vernicifera*; Copa-Psesm, copper-resistance protein A precursor from *Pseudomonas syringae*; Lac1-Neucr, laccase precursor from *Neurospora crassa*; Plas-popni, plastocyanin from *Populus nigra*; Aso-Cucma, L-ascorbate oxidase precursor from *Cucurbita maxima;* Fa8-Human 1, -2 or -3, human coagulation factor VIII precursor, domain 1, −2 or −3; Fa5-Human 1 or −2, human coagulation factor V precursor, domain 1 or −2; Ceru-Rat1, −2 or −3, caeruloplasmin precursor from *Rattus norvegicus*, domain 1, −2 or −3.

elongated molecules. Finally, sequencing of the 5′ or 3′ restriction fragments produced by cleavage with *Kpn*I, *Sal*I or *Pvu*II exclusively provided the known H1-, H2- and H3-chain sequences (Table 2, rows 3–5). This excludes the possibility of previously

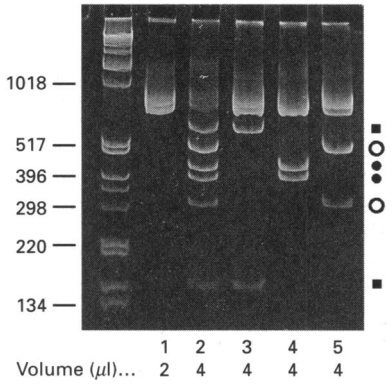


**Figure 5 RT-PCR with mouse liver RNAs and degenerate oligonucleotide primers derived from the sequences of mouse H-chain cDNAs**

The sequence of each oligonucleotide primer used for PCR is a consensus of three homologous sequences in the mouse H1-, H2- and H3-chain cDNAs: the homologous sequences for the 5′-oligonucleotide start at 877 (H1), 1041 (H2) and 854 (H3); the homologous sequences for the 3′ oligonucleotide end at 1706 (H1), 1870 (H2) or 1683 (H3). Accordingly, the degenerate oligonucleotides used were: first oligonucleotide (coding strand) 5′-TTTGT(G/T)ATTGA(C/T)(A/G)TCAG(C/T)GGCTCCATG-3′; and second oligonucleotide (complementary strand) 5′-AT(A/C/G)GTGAG(A/G)TAGGCCCA(C/G)AG-3′. The expected size of the amplified product corresponding to the H1-, H2- or H3-chain cDNA is 830 bp. After digestion with *Kpn*I, *Sal*I or *Pvu*II respectively, the expected sizes of the products are: H1, 153 + 677; H2, 435 + 395; H3, 520 + 310 (see also Figure 1). Lane 1, crude RT-PCR product derived from adult mouse liver RNAs; lane 2, same RT-PCR product restricted with (*Kpn*I + *Sal*I + *Pvu*II); lane 3, restriction with *Kpn*I only; lane 4, restriction with *Sal*I only; lane 5, restriction with *Pvu*II only. The fragments are identified on the right as follows: ■, H1; ●, H2; ○, H3. Size markers (kb) are shown on the left. The analysis was made on a 4% acrylamide gel. The relative volumes (μl) of RT-PCR products loaded in each lane were variable as indicated at the bottom.

undescribed H-chain mRNAs participating in the amplification of the *Kpn*I-, *Sal*I- or *Pvu*II-containing products. In conclusion, all our sequence data from RT-PCR products primed with degenerate oligonucleotides provide strong arguments for the existence of only three H-chain-related mRNAs in mouse liver.

### H2- and H3-chain mRNAs are transcribed in mouse brain

In mouse as in primates, transcription of the *AMBP*, *H1*, *H2* and *H3* genes takes place in liver, as judged by Northern-blot analysis [13,17,34]. We have now used the accuracy and sensitivity of our new RT-PCR assay to further investigate whether these mRNAs could be expressed in other tissues as well. Apart from trace amounts of transcripts in a limited number of tissues from adult

**Table 2 Sequencing of H-chain-related RT-PCR products from mouse liver after subcloning**

The subcloned DNA is full-length RT-PCR product (830 bp) or fragments thereof after digestion at an H-chain-specific restriction site (H1, *Kpn*I; H2, *Sal*I; H3, *Pvu*II) as indicated in Figure 1. The RT-PCR product or fragments thereof were subcloned into a pGEM-T or pUCBM20 plasmid.

| Subcloned DNA | No. of clones sequenced | Sequencing primers | Identity of the sequence read on gel |
|---|---|---|---|
| Full-length, 830 bp (apparent: 870 bp) | 4 | Universal and H2-specific oligonucleotides | H2* |
| Full-length, 830 bp | 7 | Universal | H1 or H3 |
| *Kpn*I, 3′ fragment, 677 bp | 6 | Universal | H1 |
| *Sal*I, 5′ fragment, 435 bp | 4 | Universal | H2 |
| *Pvu*II 3′ fragment, 310 bp | 6 | Universal | H3 |

* The full-length sequence of these subclones is 830 bp.

mouse (results not detailed), this study clearly demonstrated a significant expression of H2- and H3-chain mRNAs in brain. This is illustrated in Figure 6(a) where the bulk of the RT-PCR-amplified products were 830–870 bp in size (lane 1) as previously observed with liver. The amplified products could be digested with *Sal*I and *Pvu*II, i.e. the H2- and H3-chain markers respectively (lanes 4 and 5), but not with *Kpn*I (lane 3), which indicates the lack of detectable H1-chain mRNA in mouse brain. Furthermore, H3-chain mRNA clearly predominates over H2-chain mRNA in brain, as indicated by the relative signal intensities of the H2- and H3-associated bands seen before (lane 1) and after (lane 2) co-digestion with *Sal*I and *Pvu*II. That the H3-chain mRNA predominates in mouse brain was further confirmed by (i) a Southern-blot analysis of the crude RT-PCR product with H1-, H2- or H3-chain-specific probes (not shown) and (ii) a borderline detection of a single H3-chain mRNA in mouse brain by Northern-blot experiments [15].

Control RT-PCR experiments were carried out with the mouse brain RNAs or the H3-chain cDNA subcloned into a pUC plasmid as templates (Figure 6b). The brain RNA templates could still be amplified after treatment with RNAase-free DNAase I (lane 3 in Figure 6b) whereas treatment with DNAase-free RNAase prevented their amplification (lane 4). These data rule out the possibility that the H2- and H3-chain bands amplified from mouse brain RNAs by RT-PCR could result from a spurious amplification induced by minute amounts of genomic DNA.

Finally, an RT-PCR experiment with primers corresponding to mouse AMBP cDNA indicated that AMBP mRNA is not transcribed in mouse brain (Figure 6c, lane 1) whereas the corresponding RT-PCR product (517 bp) expected from liver RNAs is clearly seen (lane 2).

## DISCUSSION

To gain further insight into the structure, functions, regulation of expression and evolution of the bikunin and H-chain precursors in the IαI family, we recently began an analysis of the corresponding genes in mouse, i.e. a non-primate species [13–16]. The present study has been primarily devoted to the three H-chain precursor cDNAs. That our final sequences correspond to full-length cDNAs is indicated by the excellent match of the cDNA sizes (2904, 3125 and 2775 bp respectively) with the published sizes of mouse liver RNAs for these precursors (3.0, 3.2 and 2.9 kb respectively) [13]. Our full-length sequences in

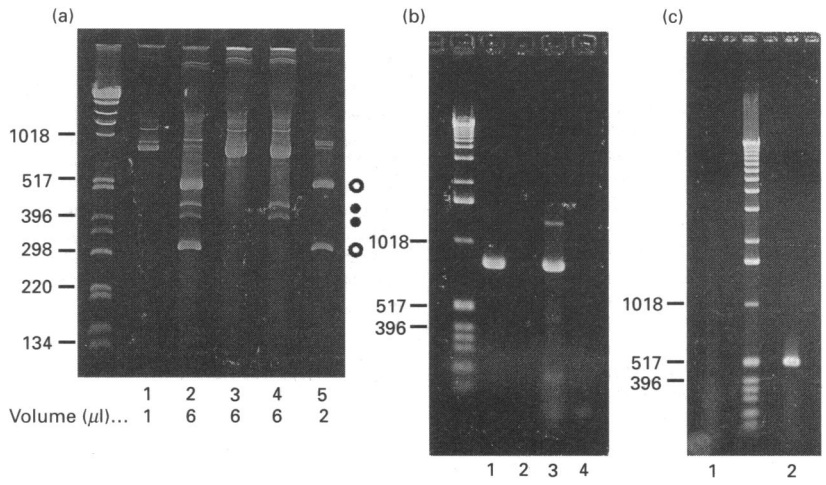


**Figure 6 Search for H-chain and AMBP mRNAs in mouse brain by RT-PCR**

(**a**) H-chain mRNAs: restriction mapping of RT-PCR products derived from adult mouse brain RNAs. The experiment, legend and symbols are the same as for Figure 5, except that liver was replaced with brain. (**b**) H-chain mRNAs: controls for RT-PCR specificity with mouse brain RNAs. Before PT-PCR, brain RNAs were treated with RNAase-free DNAase (lane 3) or DNAase-free RNAase (lane 4). To demonstrate the validity of the previous test with the RNAase-free DNAase, mouse H3-chain cDNA subcloned into pUCBM21 was left untreated (lane 1) or treated with RNAase-free DNAase (lane 2) before standard PCR. Size markers (kb) are shown on the left. The analysis was made on a 0.8% agarose gel. The volume used in each case was 20 μl. (**c**) AMBP mRNA: RT-PCR products amplified with AMBP-specific primers. The primers used were 5′-TATGTGGTCCACACCAACTATGA-3′ (coding strand) and 5′-GGACTATGGGGAGATTGCAGGCCG-3′ (complementary strand) according to the AMBP cDNA [16]. The expected PCR product is 517 bp in length. The mouse RNAs used as template were from brain (lane 1) or a control liver (lane 2). Size markers are shown in the middle. The analysis was made on a 0.8% agarose gel. The volume used in each case was 20 μl.

mouse allowed us to study extensively multiple alignments of amino acid sequences which provide new functional and evolutionary information on the IαI family, and alignments of cDNAs were the basis for transcript analysis by a newly developed RT-PCR assay.

Because the published N-terminal amino acid sequence of the human mature H2 chain obtained from purified IαI, i.e. SLPGE... [7,17], is 37 amino acids away from the signal peptide of the H2-chain precursor (see top of Figure 2), it has been proposed that a prepropeptide is encoded by H2-chain mRNA [7]. Processing of this H2-chain precursor would release the signal peptide as well as a connecting peptide (36 residues in length) located on the N-terminal side of the sequence SLPGE [7]. In keeping with the previous proposal, it has also been considered that a short propeptide in the human H1- and H3-chain precursors (H1, ATGR; H3, RSPFRLLGKR) would separate the signal peptide from the N-terminal sequence of the corresponding mature H chain [29,30], published as SKSSE... for H1 and SLPEG... for H3 [3]. However, sequence data obtained with a H1-chain polypeptide isolated from purified IαI have recently shown that the N-terminal sequence of the mature H1 chain is SATGR... [31] which contains the ATGR sequence formerly proposed as a propeptide for H1 [29]. This feature makes the presence of a propeptide sequence in the human H1-chain precursor unlikely. Likewise, the proposed limits for the signal peptide in mouse H1 chain (see the Results section) leave no room for a propeptide sequence in this chain. These features cast doubts on the overall existence of such propeptide sequences at the N-terminal side of the H-chain precursors in the IαI family. Therefore we preferred not to include any indication of such propeptide-cleavage sites in the amino acid sequences shown in Figure 2. At any rate, if a propeptide sequence did exist in the H2-chain precursor N-terminus as proposed by others [7], such a major difference between the H2-chain precursor and the H1- and H3-chain precursors would be consistent with (i) the poor sequence alignment noticed in the first 70 amino acids in Figure 2 and (ii) the marked divergence of the *H2* gene with respect to the *H1* and *H3* genes during evolution, as further discussed below.

Our amino acid similarity searches (Table 1) indicated a significant similarity of one or more H-chain precursors to the α2 subunit of a dihydropyridine-sensitive $Ca^{2+}$ channel and the α3 chain of collagen VI. These similarities are in fact accounted for by a stretch of about 200 amino acid residues, the so-called von Willebrand type-A domain [32] found in the above proteins among others, as well as in the C-terminal half of the mature H1, H2 and H3 chains (Figure 2; [35]). A heterophilic binding capacity for the proteins with one or more von Willebrand type-A domain(s) seems to be the rule [36] and the targets for proteins with von Willebrand type-A domain include such varied molecules as integrins, collagen, proteoglycans and heparin [36]. Hyaluronic acid has recently been described as a ligand for the C-terminal half of the mature H1 and H2 chains [12]. Therefore we propose that the von Willebrand type-A domains in the H chains account for the hyaluronan-liganding activity reported for some IαI family molecules [11,12,37].

The blue copper-containing oxidases and related proteins form an extremely ancient and diverse group of quite distantly related molecules found in bacteria, plants and animals. They retained or lost the ability to bind the so-called type-1, -2 or -3 copper, and may or may not have a known function related to their copper-binding domain(s) [33]. Of the plasma proteins, caeruloplasmin and coagulation factors V and VIII are examples of such molecules in which six active and/or inactive repeats of a basic copper-binding domain are indicative of domain duplications in a shared ancestor for caeruloplasmin and factors V and VIII ([33] and Figure 4). Our analyses did not reveal any sequence similarity between the H-chain precursors and caeruloplasmin or factors V and VIII, nor did we find any indication of domain duplication in the H-chain precursors. Therefore the present inclusion of the H-chain precursors in the group of proteins with multicopper oxidase domain(s) extends the number and types of plasma proteins so classified. Some members of the multicopper oxidase protein group have lost the ability to bind copper [33]. Therefore we do not yet know whether an active multicopper oxidase domain is present in the C-terminal region of the three H-chain precursors. Furthermore, this C-terminal region is trimmed off in all three mature H chains [1,7] and therefore it is not amenable to enzymic analysis carried out with purified plasma molecules such as IαI, IαLI or PαI. However, our observation of a potential metal-binding capacity for these trimmed C-terminal segments of the H-chain precursors is a first step toward the understanding of a function for these C-terminal segments.

We have investigated how many H-chain mRNAs are detectable in mouse liver with a specially designed RT-PCR assay. For this purpose, the tight homologies observed in given areas of the mouse H1-, H2- and H3-chain cDNA sequences were used to design degenerate oligonucleotides which should amplify any H-chain-related mRNA in mouse. This includes the three known H-chain mRNAs and any further H-chain mRNAs provided that they are homologous with the former three in the two areas used for oligonucleotide design. Only three H-chain mRNAs were finally detected in liver with this assay and they correspond to the known H1-, H2- and H3-chain cDNAs, as demonstrated by restriction mapping and sequencing. Therefore the existence of further H-chain mRNAs in mouse liver is quite unlikely. In man as well as in mouse, the identical level of divergence between the three H-chain precursors (Figure 3) demonstrates that the set of three H-chain genes arose from a shared ancestor before the man–mouse divergence, as recently proposed [13,14], and evolved at a very similar rate in both species thereafter. This divergence, which occurred 80 million years ago, and the 15–17% of amino acid residues that are different for any H chain between these two species (Figure 3) indicate a mutation rate of 0.2% amino acid residues per million years. Assuming that the rate of this evolution was similar before and after the man–mouse divergence, it is possible to date the separation of the *H2* and *H1/H3* genes back to 300 million years ago, whereas the *H1* and *H3* genes diverged later, about 230 million years ago. It is remarkable that this suggested timing of the occurrence of the H-chain gene family is synchronous with the proposed timing (< 270 million years ago) for the occurrence of the *AMBP* gene [16], one product of which, i.e. bikunin, is assembled with the H chains into IαI, IαLI and PαI molecules.

The dogma of a strictly liver-restricted transcription of the *AMBP, H1, H2* and *H3* genes, as inferred from Northern-blot experiments, has long been known [13,17,34]. The sensitivity and accuracy of our RT-PCR now provides unambiguous evidence for a further transcription of the *H2* and *H3* genes in mouse brain. The similar level of transcription of the three *H* genes in mouse liver contrasts with the observed predominance of the H3 transcript in brain, which suggests that the H3 chain is the major IαI-family-related polypeptide in brain. Furthermore, the observation that AMBP mRNA is undetectable by RT-PCR with mouse brain RNAs implies that the H3-chain precursor does not participate in (H3 + bikunin) assembly in brain. This contrasts with what is seen in liver where this assembly provides the PαI molecule. Likewise, the lack of transcription of the *H1* gene in mouse brain implies that synthesis of IαI molecules, i.e.

(H1 + H2 + bikunin), does not occur in brain. Therefore in brain the H3- and H2-chain precursors could be a source of single polypeptide molecules with brain-specific functions that remain to be elucidated.

We are indebted to Professor P. Arnaud for the gift of a mouse liver cDNA library. The technical assistance provided by V. Mfou'ou and C. Orvain is also appreciated. We thank Dr. L. Faye for a critical reading of the manuscript. This work was supported in part by a grant from Ministère de la Recherche et de l'Espace (attribution 91T0460), a grant from Groupement de Recherches et d'Etudes sur les Génomes to J.P.S., and the University of Rouen.

## REFERENCES

1 Gebhard, W., Hochstrasser, K., Fritz, H., Enghild, J. J., Pizzo, S. V. and Salvesen, G. (1990) Hoppe-Seyler's Z. Physiol. Chem. **371** (Suppl.), 13–22.
2 Salier, J. P. (1990) Trends Biochem. Sci. **15**, 435–439
3 Enghild, J. J., Thogersen, I. B., Pizzo, S. V. and Salvesen, G. (1989) J. Biol. Chem. **264**, 15975–15981
4 Enghild, J. J., Salvesen, G., Thogersen, I. B., Valnickova, Z., Pizzo, S. V. and Hefta, S. A. (1993) J. Biol. Chem. **268**, 8711–8716
5 Daveau, M., Rouet, P., Scotte, M., Faye, L., Hiron, M., Lebreton, J. P. and Salier, J. P. (1993) Biochem. J. **292**, 485–492
6 Kaumeyer, J. F., Polazzi, J. O. and Kotick, M. P. (1986) Nucleic Acids Res. **14**, 7839–7850
7 Gebhard, W., Schreitmuller, T., Hochstrasser, K. and Wachter, E. (1988) FEBS Lett. **229**, 63–67
8 Gebhard, W., Schreitmuller, T., Hochstrasser, K. and Wachter, E. (1989) Eur. J. Biochem. **181**, 571–576
9 Enghild, J. J., Salvesen, G., Hefta, S. A., Thogersen, I. B., Rutherfurd, S. and Pizzo, S. V. (1991) J. Biol. Chem. **266**, 747–751
10 McKeehan, W. L., Sakagami, Y., Hoshi, H. and McKeehan, K. A. (1986) J. Biol. Chem. **261**, 5378–5383
11 Chen, L., Mao, S. J. T. and Larsen, W. J. (1992) J. Biol. Chem. **267**, 12380–12386
12 Huang, L., Yoneda, M. and Kimata, K. (1993) J. Biol. Chem. **268**, 26725–26730
13 Salier, J. P., Verga, V., Doly, J., Diarra-Mehrpour, M. and Erickson, R. P. (1992) Mammal. Genome **2**, 233–239
14 Salier, J. P., Simon, D., Rouet, P., Raguenez, G., Muscatelli, F., Gebhard, W., Guenet, J. L. and Mattei, M. G. (1992) Genomics **14**, 83–88
15 Salier, J. P., Chan, P., Raguenez, G., Zwingman, T. and Erickson, R. P. (1993) Biochem. J. **296**, 85–91
16 Chan, P. and Salier, J. P. (1993) Biochim. Biophys. Acta **1174**, 195–200
17 Salier, J. P., Diarra-Mehrpour, M., Sesboue, R., Bourguignon, J., Benarous, R., Ohkubo, I., Kurachi, S., Kurachi, K. and Martin, J. P. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 8272–8276
18 Diarra-Mehrpour, M., Bourguignon, J., Sesboué, R., Mattei, M. G., Passage, E., Salier, J. P. and Martin, J. P. (1989) Eur. J. Biochem. **179**, 147–154
19 Higuchi, R. G. and Ochman, H. (1989) Nucleic Acids Res. **17**, 5865
20 Pearson, W. R. and Lipman, D. J. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 2444–2448
21 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
22 Higgins, D. G. and Sharp, P. M. (1989) Comput. Appl. Biosic. **5**, 151–153
23 Gribskov, M., McLachlan, M. and Eisenberg, D. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 4355–4358
24 Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. **12**, 387–395
25 Bairoch, A. (1992) Nucleic Acids Res. **20**, 2013–2018
26 Henikoff, S. and Henikoff, J. G. (1991) Nucleic Acids Res. **19**, 6565–6572
27 Von Heijne, G. (1986) Nucleic Acids Res. **14**, 4683–4690
28 Kozak, M. (1991) J. Biol. Chem. **266**, 19867–19870
29 Diarra-Mehrpour, M., Bourguignon, J., Bost, F., Sesboué, R., Muschio, F., Sarafan, N. and Martin, J. P. (1992) Biochim. Biophys. Acta **1132**, 114–118
30 Bourguignon, J., Diarra-Mehrpour, M., Thiberville, L., Bost, F., Sesboué, R. and Martin, J. P. (1993) Eur. J. Biochem. **212**, 771–776
31 Malki, N., Balduyck, M., Maes, P., Capon, C., Mizon, C., Han, K. K., Tartar, A., Fournet, B. and Mizon, J. (1992) Biol. Chem. Hoppe-Seyler **373**, 1009–1018
32 Sadler, J. E. (1991) J. Biol. Chem. **266**, 22777–22780
33 Ryden, L. G. and Hunt, L. T. (1993) J. Mol. Evol. **36**, 41–66
34 Bost, F., Bourguignon, J., Martin, J. P., Sesboué, R., Thiberville, L. and Diarra-Mehrpour, M. (1993) Eur. J. Biochem. **218**, 283–291
35 Bork, P. and Rohde, K. (1991) Biochem. J. **279**, 908–910
36 Colombatti, A. and Bonaldo, P. (1991) Blood **77**, 2305–2315
37 Hutadilok, N., Ghosh, P. and Brooks, P. M. (1988) Ann. Rheum. Dis. **47**, 377–385

Received 21 July 1994/5 October 1994; accepted 2 November 1994